基于机构知识库的查收查引功能设计与实现*

■ 刘艳民[1] 祝忠明[2] 张旺强[2]
[1] 兰州大学图书馆 兰州 730000 [2] 中国科学院兰州文献情报中心 兰州 730000

摘要：[目的/意义]为提高引证检索服务效率，方便科研人员自助查询收录引证报告，扩展机构知识库提供的科研评价服务功能，开发基于机构知识库CSpace系统的查收查引功能。[方法/过程]对人工查收查引服务的关键流程和细节问题进行调研，提出机构知识库扩展查收查引功能的必要性，设计功能实现流程图，实现科研成果显示、收录类型和引用次数、他引次数统计、施引文献和SCI、EI、CSCD等详细信息导出功能。[结果/结论]基于机构知识库的查收查引功能经过测试和评估，可以达到95%以上的收录引用准确率，相比人工检索，服务效率提高明显，可以帮助图书馆节省人力资源约75%。真正实现机构知识库对查收查引服务的有效支撑。

关键词：查收查引 机构知识库 功能设计 引证检索
分类号：G250.7
DOI：10.13266/j.issn.0252-3116.2018.12.012

1 引言

查收查引是检索机构根据用户需求，在国内外权威数据库中检索其论文被收录和引用情况，以证明其科研能力和水平而开展的信息咨询服务。具体来说，是通过作者姓名、单位、期刊名称及卷期、会议信息、篇名等途径，查找论文被SCI(Science Citation Index)、SSCI(Social Science Citation Index)、A&HCI(Arts Humanities Citaion Index)、EI(The Engineering Index)、CPCI(Conference Proceedings Citation Index)、中国社会科学引文索引数据库(Chinese Social Sciences Citation Index, CSSCI)、中国科学引文数据库(Chinese Science Citation Database, CSCD)等权威数据库收录及被引用情况，并根据检索结果出具检索证明报告。收录引证报告作为课题申报、职称评定等的重要参考依据。

通过笔者调查发现，目前大多数图书馆开展的引证检索服务，是通过委托人提供查收查引委托单，工作人员手动检索指定的数据库查询委托人提供的论文，并对下载数据进行人工整理与统计，最后形成引证报告。以兰州大学图书馆为例，2013-2016年平均每年做引证检索报告500份，每份报告少则花费几小时完成，多则一天甚至两天才能完成，尤其在检索高峰期，则需要更多的时间。这项工作重复性高、效率较低、花费大量的劳力和时间，即无法满足用户的全部需求，也影响服务质量。因此，利用计算机程序在线完成查收查引流程中的论文清单提交、论文被收录、引用及他引次数统计、引证报告自动导出等工作成为查收查引工作的必然发展趋势。

2 引证检索服务现状与机构知识库扩展查收查引功能的必要性

2.1 引证检索服务现状

查收查引是我国图书馆独具特色的服务项目，人工引证检索服务主要由10个步骤组成[1]，分别为：委托单接受和确认、论文清单核查、检索任务分工、论文收录检索、论文引用检索、检索结果格式整理、检索结果统计、检索报告撰写、用户核查报告、费用支付和报告领取。该服务步骤繁琐、效率较低，且工作过程中存在多库检索耗时耗力、检索结果整理繁琐、检索结果不能重复使用、检索报告不能自动生成等主要问题[2]。

在机器辅助引证检索服务方面，查新机构做出了大量的实践研究。清华大学最早研发了代查代检服务

* 本文系中国科学院文献情报能力建设专项"机构知识库持续运行建设及开放科研知识云"(项目编号：Y7ZG011001)研究成果之一。
作者简介：刘艳民(ORCID:0000-0001-7054-3498)，馆员，硕士，E-mail:liuyanmin@lzu.edu.cn；祝忠明(ORCID:0000-0002-2365-3050)，研究员，博士；张旺强(ORCID:0000-0002-5105-598X)，馆员，硕士。
收稿日期:2018-01-30 修回日期:2018-04-23 本文起止页码:91-97 本文责任编辑:徐健

系统[3]。北京大学图书馆的李晓东、卢振波 2005 年提出通过工具软件实现作者论文数据的采集、自动检索和自动下载等功能，但未对查收查引系统进一步设计与实践[4]。樊亚芳等提出利用 Excel 的筛选功能和 EndNote Web、NoteEpress 等文献管理软件辅助手工检索和实现对自引文献的批量去除、他引文献列表的格式化输出、以及总他引次数、引用期刊种数和作者人数的统计[5-7]，该方法可辅助人工检索提高服务效率。在查收查引业务流程自动化管理方面，山东大学的师晓青[8,9]设计并构建了基于 B/S 架构的高效图书馆查收查引检索系统，详细规划和设计了 6 个角色用户的权限和主要功能，用户可以在线提交检索申请书，检索员可对报告进行管理，审核员可对报告审核、归档保存。徐世妍[10]提出引入 JBPM 工作流技术设计全新的查收查引综合服务平台，这两系统的优势是流程自动化，不足之处是查收查引的关键检索功能、报告生成等未实现自动化。北京邮电大学图书馆的严潮斌、陈嘉勇[11,12]等提出了将查收查引服务融入机构知识库生态圈的观点，以查收查引为灵感，自主研发机构知识库，实现文献与作者之间的精准关联机制，能够精准关联出作者被收录和引用的文献列表，但在查收查引工作的数据问题方面存在不准确性，需要到数据库中实际验证，且没有实现引证报告自助导出等功能。

CALIS 技术中心与北京大学图书馆[1]针对论文收录及引用联合开发了一款 CALIS 论文收录及引用检索系统。该系统基本完成了手工查收引证 10 个步骤中的 7 个，除了委托单的接受和确认、论文清单初步检查和用户核查报告仍需要传统方式。该系统经过 6 轮测试和近半年的试用改进，基本满足高校图书馆的需求，极大缓解了人力资源紧张。目前已有 150 多家大学图书馆开通试用，包括 7 所“985”高校，正式购买用户 19 家。但论文清单提交和接收需要手工完成，检索报告没有期刊影响因子，没有统计委托人为第一作者的文献数等更多信息。2011 年底，中国科学院软件研究所研发了“引证报告自动生成原型系统”，王学勤[13-14]等在原型系统的基础上进行优化，增强了数据预处理功能和算法，增加了检索数据源、人机交互模块及自引排除等功能模块。该系统经测试评估工作效率、准确率、稳定性都达到了预期的目标，且已在中国科学院文献情报中心部署使用，效果整体良好。但仍需完善用户并发控制、用户权限管理、中文及会议论文引用检索、生成多格式报告等功能。

针对已有查收查引系统存在需要手动提供委托单、缺乏期刊影响因子等评价指标、对中文检索效果不佳等问题，笔者开发了基于机构知识库(Institutional Repository，IR)的查收查引功能模块。选择具有代表性的 CALIS 收录引用检索系统、中国科学院(以下简称“中科院”)开发的引证报告自动生成系统与笔者开发的 IR 查收查引功能模块从服务模式、开发方式、检索效果、实现功能及应用效果等方面进行比较分析，见表 1。

2.2 IR 扩展查收查引功能的必要性

IR 作为很多高校和科研机构保存管理科研知识成果的重要手段，经过多年的建设和维护，积累了一定的学术资源。如何提升 IR 系统对高校和科研机构的科研成果评价服务成为未来发展面临的重要课题。下面列出 IR 扩展查收查引功能的必要性和有利条件。

首先，IR 中存缴了本机构作者的大量学术成果的元数据，为查收查引奠定了良好的数据基础。在建设初期，元数据主要通过专业平台批量导入，英文数据库包括 Web of Science、EI 等收录的期刊论文和会议论文，中文数据库包括 CNKI、CSCD、CSSCI。IR 也通过 Science Router 提供学术资源自动采集服务，支持从大型主流学术资源库中采集机构公开发表成果，并以接口方式共享数据。学术资源库包括 IEEE、Springer、PubMed、ELSEVIER、Google Scholor、CiteSeerx 等。且自动从 SciRouter 获取本机构最新产出数据，同步更新批量导入数据库。由于 IR 存缴的论文覆盖面广，基本涵盖了本机构所有学术成果。查收查引检索的科研成果其实是 IR 的子集，对查收引用户来说，无需提供待检索文献。IR 对导入的元数据进行规范处理，解决了用户提交论文清单出现的列表格式不一、数据不全等问题。

其次，IR 的作品认领机制实现了作者与科研成果的无缝匹配。IR 平台通过建立作者别名数据库和作者唯一标识符[15]，采用机器的自动方式匹配出作品与作者之间可能的关联，将关联信息推送给相关作者进行认领，并保存认领结果，实现了作者信息和相关作品信息之间的准确关联。在人工查收查引过程中，需要根据论文清单，按照以文献标题为主，结合作者、期刊名称、会议信息，在指定数据库进行收录查证，对用户提交的待查文献按“首作者 & 被引著作”进行引用检索，并自动对检索结果进行错引确认，这个过程步骤繁琐，效率低。IR 中可以实现用户自助查询个人所有学术成果，引证检索人员也可避免重复检索多个数据库。

再次，IR 与查收查引工作相辅相成，互相促进，且

表1　查收查引典型系统分析

比较内容		CALIS 查收查引系统	中科院引证报告自动生成系统	IR 查收查引功能模块
服务模式		在线试用/本地部署	B/S	B/S
开发方式		自行构建	自行构建	基于 CSpace 系统
功能模块	文献预处理	手动著录和清单导入（包括委托人清单、模板清单和历史清单）	以 Word、Excel、BibTex、EndNote 等格式上传文献，支持在线检索，规范待查文献	用户不需提交文献清单，IR 构建的数据、作者认领作品形成文献列表，文献列表字段可扩展
	收录检索	操作流畅，可以进行原始信息和检索信息的对比。可以检索常规数据库，包括 EI、WOS、CSCD 等	以文献标题从数据库提取数据，存储在系统自建 Mysql 数据库进行检索。可以检索常规数据库，包括 EI、WOS、CSCD 等。对中文数据库检索效果一般	收录检索在导入 IR 过程中进行，以 IR 中文献对应的收录字段进行统计。收录类型全面，包括 Pubmed、Medline、Biosis 等较多类型
	引用检索	提供疑似被引文献记录供人工确认，自动区分他引/自引，对作者难以正确区分的进行人工干预判断	以作者 & 刊物或作者 & 页码从 SCI、CSCD 提取数据，进行人工错引、自引确认	通过 SCI、CSCD 的接口程序获得被引、施引文献采集存储在 CSpace 系统，进行自引排除、他引自动化计算
	报告生成	生成多种格式的报告，提供对引用信息的“全引”和“只要自引”的选择	报告格式单一	根据模板自动生成，用户可根据需求定制模板
	计费	提供校内转帐等形式的计费	无	无
检索效果评估	英文数据库	原文规范，系统可自动匹配收录 97% 以上，引用系统匹配率较低 7.42%，需人工确认匹配 92%。原文不规范时系统会自动匹配收录 80% 左右，对查引用正确率系统匹配 45%	SCI 论文的收录和引用准确率达到 90% 以上，收录〉90%，引用〉95%	收录引用准确率 95%
	中文数据库	原文规范，收录正确率 100%，引用正确率系统匹配达到 62%。原文不规范，收录正确率 94%，引用系统匹配 90%，剩余需人工确认匹配	中文数据库未评估	收录引用准确率 98%
应用效果		支持多进程、多任务并行，可同时处理多个数据库的检索，允许中英文文献同步检索，检索总耗时为手工的 1/3。150 多家大学图书馆开通试用，包括 7 所 985 高校，19 所高校正式购买使用。	经过 3 次系统改版和两次引证查询高峰期的检验。安装在中国科学院文献情报中心，每年出具检索报告 750 份。收录引用准确率达到 90% 以上，对 50 篇以上文献，最大工作效率达 260%，不支持多用户并发控制、用户权限管理。	支持多用户，多进程同时处理，节省图书馆人力资源 75%。110 多家研究所和 20 多所高校图书馆（安装部署且升级到机构知识库 CSpace6.0 系统）

拥有强大的数据导出功能。高校图书馆查收查引服务用户群基本固定，会存在委托人几年内连续多次委托查询，或者同一用户更新论文清单的收录引用情况。由于图书馆手工操作流程的弊端，报告更新相当于重新检索，不但增加了重复检索的费用，且不能满足用户快速获得报告的需求。IR 中的查收查引功能较好解决图书馆和用户面临的这些问题。对于 IR 建设者来说，查收查引不再是一次性的重复劳动，通过查收查引工作，可以补充 IR 中暂时没有收录的作品，督促作者和 IR 建设者来更新数据。

最后，IR 可促进查收查引服务综合多样化。查收查引作为科技评价的重要手段之一，常用的指标为收录篇数、引用次数及他引次数。为使查收查引由简单评价向综合评价发展，需要评价指标的综合多维化，IR 提供的学科分析支持服务，评价指标较全面。其中，科研成果有是否为 A 类论文、ESI 高被引论文、高热点论文等；收录类型包括 SCIE、SSCI、A&HCI、PUBMED、MEDLINE、BIOSIS、CPCI－S、ESI、CPCI－SSH、CSSCI、CSCD、EI 等。IR 的论文发表期刊分布分析中，期刊的配置项有 JCR 分区、CAS 分区、JCR 影响因子、期刊类型。这些元数据为促进查收查引服务的综合多样化奠定了良好的基础。

上述因素构成 IR 中扩展查收查引功能的重要依据。

3　IR 中查收查引功能框架与实现

3.1　设计思路及功能框架

笔者基于中科院 IR 系统已有的资源及其存储结构，结合目前查收查引服务的自动化需求，设计开发了 IR 中的自助导出收录引证报告功能，旨在为科研成果评价提供有效支撑。IR 的收割、存缴论文机制、作品认领机制，可实现显示用户论文清单列表。提取查收查引中科研成果需要的主要字段，可以根据用户需求扩展字段。IR 中论文的收录类别字段是提交期刊论文和会议论文原数据信息时默认采集的，在资源建设的过程中对该原数据的采集重视度较高，数据质量较好，且 IR 对一篇文献被多个数据库重复收录的情况进行了查重处理，在条目导入数据库的时候，首先进行题

名 + 作者匹配查重，如果条目重复，则添加相应的收录类型到收录类别字段。例如，兰州大学 IR 中有一篇论文同时被 SCIE、EI、PubMed、Medline、BIOSIS、CSCD 数据库收录(见图 1)，收录类别超出查收需求。如手工去检索，需要检索 5 个数据库平台，工作人员因为只习惯使用一两种检索式进行检索，工作量倍增，且容易发生漏检的情况。IR 通过查重匹配，收录类型全面，既提高了查准率，也提高了查全率，保证查收功能自动化实现的可行性。

IR 查收查引功能框架，该功能由两个模块组成，即作者文献列表显示模块和导出引证报告模块，这两个模块基本实现了数据预处理、收录引用检索统计、报告生成等功能。且都依赖于 IR 系统的完整数据结构。IR 底层的数据来源包括了 WOS、EI、CSSCI、CSCD、CNKI 等数据库平台，以及 Web of Science、CSCD 提供的被引频次接口程序。见图 2。

3.2 关键功能实现

IR 中查收查引功能模块实现了作者文献列表即数据预处理功能、论文收录引证检索功能、施引文献采集及他引/自引区分功能、引证报告导出功能。作者论文清单列表见图 3。引证统计表包含了 JCR 影响因子、SCI 被引次数、CSCD 被引次数、及论文题名、作者、来源期刊、发表时间等主要字段。

人工统计收录总篇数、总引用次数的步骤主要是：通过从数据库导出要查询的论文，摘取相应的字段，填入检索结果表单中，

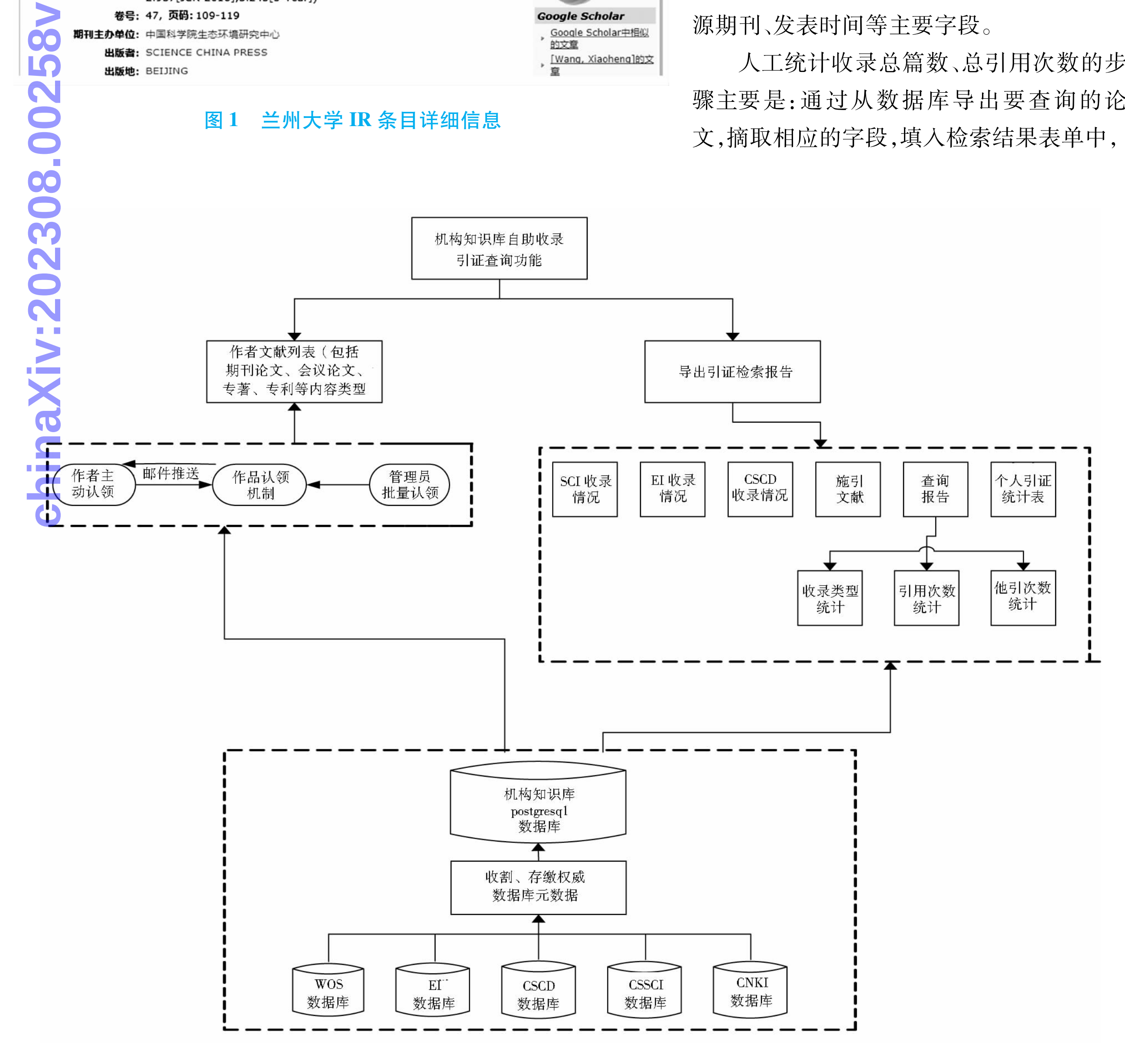


图 1 兰州大学 IR 条目详细信息

图 2 IR 中查收查引功能框架

图3 IR中作者个人作品引证统计表

借助EXCEL得出总被引次数。IR统计总被引次数，只需提交文献，后台程序自动计算总被引，有较高的效率和准确率。

（1）施引文献信息采集处理功能实现。引证检索报告中的施引文献数据量较大，如文献的被引次数小于100，则检索和导出施引文献相对容易。但对个别文献被引次数达到几千次，如兰州大学物理科学与技术学院耿柏松教师有一篇论文被引次数达到1 144次，由于WOS平台的条数限制（每次最多标记50条），手动下载需要翻页23次。IR导出引证检索报告实现了对施引文献的自动获取，通过对IR数据库中条目的元数据字段wos_citing_url（施引文献URL）进行分析，运用网页信息提取技术，即从HTML页面找到描述文献的网页元素，并提取出相应的数据内容。采用Jquery技术，获得所有施引文献总数及元数据字段，包括题名、作者、WOS记录号、来源期刊等信息。为解决导出施引文献速度较慢，笔者设计了定时任务，将URL抓取到的施引文献信息存入IR自建数据库，大大提高了导出施引文献效率。

（2）他引和自引区分功能实现。在文献引证检索服务工作中，他引被用作科技成果影响力评价的主要指标，通过排除论文自引而获得。自引排除方法一般有排除一个作者既被检索作者、排除团体作者、排除论文全部作者。本研究中采用的他引计算方法为严格他引，既排除论文全部作者。IR对论文的施引文献进行采集，存储的字段包括作者、作者单位等信息。且施引文献与原文献都来自Web of science数据库平台，作者是用全名匹配，对个别全名不规范的作者，在数据库中进行规范处理。他引次数计算实现流程见图4。

4 查收查引功能测试和效果评估

IR系统的查收查引功能实现以后，笔者从时效性、稳定性、收录引用正确率3个方面对该功能进行了测试和评估。选取的测试数据是兰州大学IR已建好

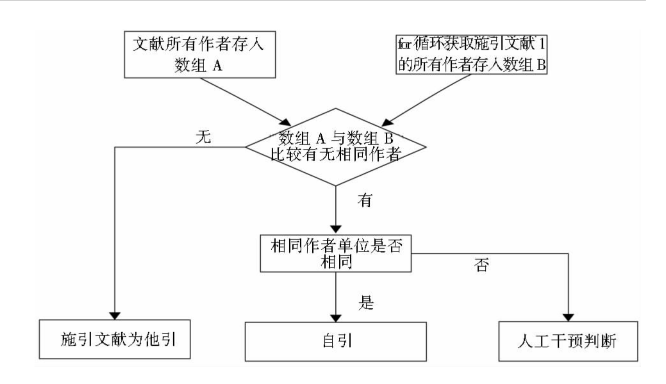


图4 他引次数计算流程

的资源，涵盖学科有物理、化学、生物、经济等重点学科。测试论文总篇数12 524篇，其中被SCIE、EI共同收录的论文7 216篇、CSCD收录论文3 805篇、CSSCI收录论文1 259篇、CPCI-S收录论文234篇，SSCI收录论文10篇。涉及兰州大学学者356位。其中，SCIE收录论文被引次数最高为1 168，CSCD被引频次最高为252。对中文数据库收录的论文，即被CSCD和CSSCI收录的论文，测试系统用时和收录引用正确率如表2所示：

表2 中文文献导出报告用时及准确率

文章数（篇）	人工用时（分钟）	系统用时（统计收录、引用、导出报告）(s)	系统收录引用正确率（%）
20以内	5－20	<5	>99
20－50	70－90	<10	>99
50－100	120－300	<15	>98
100－150	300－450	<30	>98

对于中文文献，只需统计收录类型、引用次数，导出报告只需导出CSCD及CSSCI收录详细信息。因此，在IR中速度较快，收录详细信息来自机构库数据库，引用次数通过接口获得。对于100篇的文献，导出检索报告只需30s左右，相比人工检索，节省时间大约在5－7小时，与CALIS查收查引系统相比，节省时间大约1－2小时。收录引用准确率达到98%，这归功于在IR建设的过程中，提交作品的工作者对数据的严格把控和规范。而CALIS查收查引系统对中文文献的收录引用准确率在原文规范的情况下很高，而在原文不规范时系统自动匹配准确率只有90%左右，查引用准确率低于收录准确率，大约10%左右的文献需人工确认匹配。

在英文文献导出报告时，则需要花费较多时间，对SCIE收录论文需计算他引次数及导出施引文献，英文文献导出报告花费时间及准确率如表3所示：

表 3 英文文献导出报告用时及准确率

总被引次数	人工用时（分钟）	系统用时（统计收录、引用、他引、导出施引文献等）（S）	系统收录引用正确率（%）
<10	5 – 10	8 – 10	>97
10 – 50	30 – 180	10 – 45	>96
50 – 200	200 – 750	60 – 180	>95
200 – 500	750 – 1 500	180 – 300	>95

从表 3 中可以看出，英文文献在计算他引次数和导出施引文献时，花费时间相对中文较多，随着总被引次数的增高，花费时间也在增长，对被引次数在 500 左右时，导出报告花费时间在 5 分钟左右，而人工整理需要 30 个小时左右，中科院引证报告自动生成系统花费时间大约在 7 – 8 小时。对被引次数最高达 1 168 的论文，系统测试导出报告仅需 10 分钟左右。而对收录引用准确率随着被引次数的增多，会有所下降，因英文数据库会出现错引，需人工核对，施引文献越多，出错的可能性越大。在测试 12 524 篇论文的收录引证报告过程中，系统运行稳定。CALIS 查收查引系统对英文数据库的检索准确率主要依赖于论文清单是否正确规范，对于信息有误的论文清单或缺失信息较多时，英文文献查收录正确率系统自动匹配 80% 左右，17% 左右需人工确认匹配，错误检索 2% 左右，查引用正确率 50% 需人工确认匹配。

5 结论

IR 的查收查引功能跟手工检索相比，大大缩短了检索时间，提高了检索效率，节省大量劳动力。且可以达到 95% 的收录引用准确率，能够满足高校委托高峰期用户检索需求。与 CALIS 查收查引系统和中科院引证检索系统相比，精简了提交委托单以及查收引人员检索多个数据库的过程，用户只需登录机构库，查询需要检索的论文，点击提交就可完成导出报告功能。机构知识库的查收查引功能未来拓展空间较广，用户可以根据需要导出需要的期刊分区、H 指数等指标。为科研评价的综合多维化提供可能性。

本文开发的查收查引功能完全依赖于 IR，因此需进一步完善解决 IR 中资源的数量和元数据质量问题。针对 IR 数据质量存在的问题，需从源头规避，IR 构建者在检索数据库时，应根据专业检索人员的指导，采用多种检索方式去检索数据库，以防漏检，规避由于不规范引用造成施引文献检索结果的不准确；对导入 IR 的数据进行规范处理，可及时更正数据库中的错误信息；机构用户认领作品信息时对文献详细信息进行及时核对，对错误信息修改更新；对资源的实时更新，通过 iSwitch 接口提高资源存缴效率和完善元数据质量。系统需进一步完善查收查引业务流程自动化功能，以更好地为科研评价提供服务。望笔者开发的查收引功能在高校图书馆得到广泛应用，也为 IR 的可持续建设提供支撑。

参考文献：

[1] 马芳珍，李峰，季梵，等. 对 CALIS 查收查引系统的测试和应用效果评价[J]. 大学图书馆学报，2016，34(2)：97 – 102.
[2] 马芳珍. 查收查引系统需求分析和设计要点探讨[J]. 大学图书馆学报，2015，33(4)：80 – 84，121.
[3] 战玉华，程爱平，钱俊雯，等. 代检代查服务系统的开发及应用[J]. 图书情报工作，2005(11)：75 – 77，55.
[4] 李晓东，卢振波. 论文查收查引工具软件的设计与实现[J]. 大学图书馆学报，2005(1)：49 – 50，62.
[5] 樊亚芳，陈锴. 利用 Excel 和 EndNote Web 提高论文查收查引工作效率[J]. 图书馆杂志，2013，32(1)：32 – 34，60.
[6] 张雪娟，樊亚芳. NoteExpress 在论文查收查引工作中的应用[J]. 情报探索，2017(6)：45 – 49.
[7] 樊亚芳. 利用文献管理软件提高论文查收查引工作效率的实践与应用[J]. 高校图书馆工作，2017，37(2)：63 – 66.
[8] 师晓青，尉江星. 基于 JSP 的在线查收、查引和查新检索系统的设计与实现——以山东大学(威海)为例[J]. 现代情报，2014，34(3)：131 – 134，138.
[9] 师晓青. 高校图书馆查收查引系统设计与实现[D]. 济南：山东大学，2014.
[10] 徐世妍. 基于 jBPM 的查收查引综合服务平台设计[J]. 新世纪图书馆，2015(11)：52 – 56.
[11] 侯瑞芳，陈嘉勇，周婕. 查收查引服务优化体系的构建与思考[J]. 图书馆建设，2015(4)：75 – 79.
[12] 严潮斌，陈嘉勇，侯瑞芳，等. 查收查引服务支撑需求驱动下的高校机构知识库建设[J]. 现代图书情报技术，2015(5)：94 – 100.
[13] 王学勤，郝丹，郑菲，等. “查收查引报告自动生成系统”应用实践研究[J]. 图书情报工作，2014，58(16)：131 – 137.
[14] 郝丹. 引证检索中数据质量控制研究与实现[D]. 西安：西安电子科技大学，2012.
[15] 刘巍，祝忠明，张旺强，等. 机构知识库中作者标识与作品认领机制的研究与实现[J]. 现代图书情报技术，2014(3)：8 – 13.

作者贡献说明：

刘艳民：程序开发和论文撰写；
祝忠明：提出研究思路和指导论文撰写；
张旺强：程序开发和功能测试。

刘艳民，祝忠明，张旺强. 基于机构知识库的查收查引功能设计与实现[J]. 图书情报工作,2018,62(12):91 - 97.

Design and Implementation of Cited Reference Retrieve Function Based on Institutional Repository

Liu Yanmin[1] Zhu Zhongming[2] Zhang Wangqiang[2]
[1] Lanzhou University Library, Lanzhou 730000
[2] Lanzhou Library of Chinese Academy of Sciences, Lanzhou 730000

Abstract: [**Purpose/significance**] The paper aims to improve the efficiency of citation retrieval service, facilitate researchers to query citation report by self-service, expand the function of scientific research evaluation service of institutional repository, and develop the cited reference retrieve function of CSpace system based on institutional repository. [**Method/process**] It researches the key process and details of artificial service, proposes the necessity of IR to extend functionality, designs the flow chart of function realize, and uses the J2EE framework to realize the function. [**Result/conclusion**] The function has been tested and evaluated, which can achieve more than 95% of the citation accuracy. Compared with manual retrieval, the service efficiency is improved obviously, and the library saves about 75% of human resources. The institutional repository has realized the effective support for the investigation and retrieval service.
Keywords: cited reference retrieve service institutional repository function design cited reference retrieve